Information Fusion 44 (2018): 57-64

DOI: https://doi.org/10.1016/j.inffus.2017.12.003

Multiple classifiers in biometrics.
Part 1: Fundamentals and review

Julian Fierrez, Aythami Morales, Ruben Vera-Rodriguez, David Camacho

School of Engineering, Universidad Autonoma de Madrid, Madrid, Spain

---

Abstract

We provide an introduction to Multiple Classifier Systems (MCS) including basic nomenclature and describing key elements: classifier dependencies, type of classifier outputs, aggregation procedures, architecture, and types of methods. This introduction complements other existing overviews of MCS, as here we also review the most prevalent theoretical framework for MCS and discuss theoretical developments related to MCS.

The introduction to MCS is then followed by a review of the application of MCS to the particular field of multimodal biometric person authentication in the last 25 years, as a prototypical area in which MCS has resulted in important achievements. This review includes general descriptions of successful MCS methods and architectures in order to facilitate the export of them to other information fusion problems.

Based on the theory and framework introduced here, in the companion paper we then develop in more technical detail recent trends and developments in MCS from multimodal biometrics that incorporate context information in an adaptive way. These new MCS architectures exploit input quality measures and pattern-specific particularities that move apart from general population statistics, resulting in robust multimodal biometric systems. Similarly as in the present paper, methods in the companion paper are introduced in a general way so they can be applied to other information fusion problems as well. Finally, also in the companion paper, we discuss open challenges in biometrics and the role of MCS to advance them.

Keywords:
classifier, fusion, biometrics, multimodal, adaptive, context

---

1. Introduction

The basic aim of pattern recognition is to devise automatic procedures that maximize certain criteria for the recognition problem at hand, usually related to the recognition performance. This is normally achieved by comparing different existing pattern recognition algorithms on the specific problem studied, and selecting the best of them. Worth noting, by observing the errors misclassified by the different approaches, one can observe that some recognition errors committed by the best approach can be well resolved by the inferior methods. These observations motivated a big interest in combining classifiers in the 90's [1], which was followed by very active research since then. This is exemplified by the successful series of Workshops on Multiple Classifier Systems (MCS), conducted yearly since 2000 [2, 3].

This multiple classifier approach can be found with different names in the literature [4]: classifier combination, classifier fusion, mixture of experts, committees of neural networks, consensus aggregation, expert conciliation, voting pool of classifiers, dynamic classifier selection, composite classifier design, classifier ensembles, divide-and-conquer classifiers, etc.

In addition to important theoretical advances, the above mentioned research in multiple classifier systems has resulted in highly successful practical developments in almost any field in which pattern classifiers are used, e.g., analytics of data streams [5], astronomy [6], biometrics person recognition [7], computer vision and medical image analysis [8], decision making [9], document analysis [10], hybrid systems [11], machine learning [12], neural information processing [13], and many others. One prototypical example of a big practical MCS success is the Viola-Jones cascade classifier [14], one of the most cited and widely used approaches in computer vision.

In the previous paragraph and related literature [4], the reader can find excellent surveys of MCS methods and algorithms. Out of those previous general references, the most related publications are the excellent MCS overview by Polikar [9], which is still a valuable reference after more than 10 years, and the quite recent overview of MCS applied to biometrics by Lumini and Nanni [15]. We complement the first overview by Polikar being more general, up to date, and more focused into fundamentals. On the other hand, in [9] one can find introductory descriptions of specific MCS algorithms like the ones only mentioned here in Table 2. With regard to the recent overview by Lumini and Nanni [15], here we are more comprehensive in our review of methods for biometrics, including the basics of the most prevalent theory of

MCS applied to biometrics. We also develop recent trends and developments not discussed by Lumini and Nanni, including adaptive architectures and practical algorithms implementing the discussed new trends.

In order to be as self-contained as possible while avoiding overlap with related publications, this paper is divided into two Parts, each of them with slightly different intended audience.

In the present paper, Part 1, we first provide a brief introduction to MCS outlining basic nomenclature, architecture, and key elements, with a focus into the fundamentals of MCS. We refer the reader to the references in previous paragraphs for descriptions of established MCS methods and algorithms.

After the brief introduction to MCS, we also review here in Part 1 the application of MCS to the particular field of multimodal biometric person authentication in the last 25 years or so, as a prototypical area in which MCS has resulted in important achievements. We review MCS in multimodal biometrics with general descriptions of main MCS elements, methods, and algorithms; facilitating the export of experiences and methods to other information fusion problems.

The companion paper, Part 2 of this series of two papers, is intended for researchers knowledgeable in MCS interested in recent developments in context-based information fusion coming from the biometrics research community, or newcomers to MCS that have first addressed Part 1.

Based on the theory and framework introduced in Part 1, in the companion paper, Part 2, we develop in more technical detail recent trends and developments in MCS from multimodal biometrics that incorporate context information in an adaptive way, in the form of input quality measures [16], and user specific particularities that move apart from general population statistics [17]. These MCS architectures exploiting context-information have been highly successful in multimodal biometrics, and may find good application in other information fusion fields as well. In related works such as [18], one can find an excellent treatment of general context-based information fusion, in which the methods and specific algorithms developed in Part 2 can be well accommodated and applied.

The companion paper ends with a discussion of open challenges in biometrics that can be addressed and advanced using MCS. The challenges exposed largely follow the excellent survey and outlook of the field of biometric person recognition by Jain et al. [19], which we complement with our personal view, and augment with the way MCS developments can advance key challenges in biometrics.

Biometrics person recognition shares many issues and challenges with other pattern recognition applications like video surveillance [20], speech technologies [21], human-computer interaction [22], data analytics applications [23], or recommender systems [24]. The last section here in Part 2 focus on ways how MCS may advance challenges in biometrics, but looking at that, the reader may also find useful paths to the future of other pattern recognition and information fusion areas as well.

2. Multiple Classifier Systems (MCS)

Multiple classifier approaches can be categorized depending on: assumption about classifier dependencies, type of classifier outputs, aggregation procedure, and architecture.

Classifier dependencies. In general, we may have different classifier outputs because of [25]: different feature sets, different training sets, different classification methods, different parameters in the classification method, or different training sessions. All these reasons result in a set of classifiers whose outputs may be combined with the hope of improving the overall classification accuracy. Classifier combination is specially useful if the individual classifiers are largely diverse [26]. If this has not been guaranteed by the use of different training sets, resampling techniques like rotation or bootstrap may be used to artificially create such differences. Examples of classifier combination based on resampling strategies are the well known stacking [27], bagging [28], and boosting [29].

In the case of multimodal biometric authentication, the independence between classifiers (one for each modality) is normally assumed.

Type of classifier outputs. The outputs of the different classifiers can be classified into three levels [30]: 1) abstract, 2) rank, and 3) measurement (or confidence). At abstract level, each classifier only outputs a class label. At rank level, each classifier outputs a ranked list of classes, with the class ranked first being the first choice. At measurement level, each classifier outputs a numerical value indicating the belief or probability that the pattern belongs to a given class.

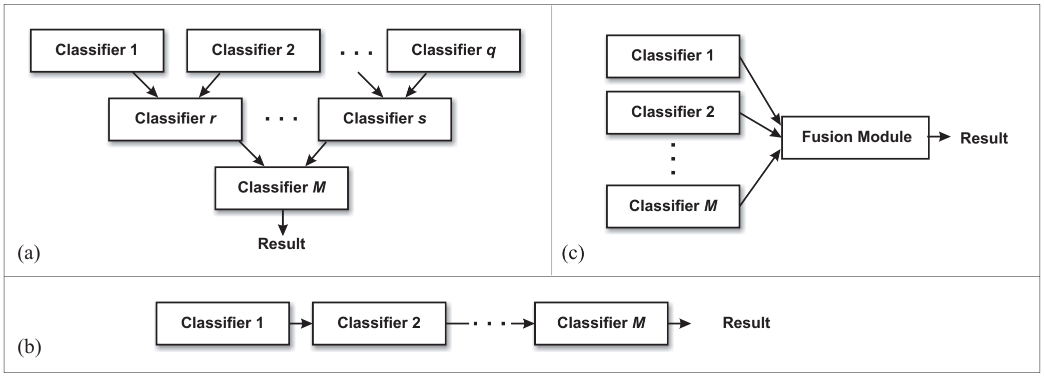

Figure 1: Architectures for multiple classifier combination: (a) hierarchical, (b) serial, (c) parallel.

Aggregation procedures. Aggregation procedures can be first classified according to trainability and adaptivity. Some combiners do not require training while others are trainable. The trained combiners may lead to better performance at the cost of additional training data and additional training. Some combiners are adaptive in the sense of weighting the contribution of each expert depending on the input pattern. Conversely, nonadaptive combiners consider all input patterns in the same way. Adaptive schemes can exploit the detailed error characteristics of the individual classifiers under different input patterns. Examples of adaptive combination strategies include adaptive weighting [31], mixture of local experts (MLE) [32], and hierarchical MLE [33].

Architecture. The schemes for multiple classifier combination can also be grouped according to their architecture into three main categories [25]: 1) hierarchical (or tree-like), 2) cascading (or serial), and 3) parallel. A graphic representation of the three categories is given in Fig. 1.

In hierarchical classifier combination schemes, the different classifiers are combined into a tree-like structure. This is the more flexible architecture and enables to exploit the different discriminative power that can be embedded in different groups of features.

In the cascade architecture the classifiers are invoked in sequence. Some of them may only be used if certain conditions occur in the outputs of the classifiers invoked first. This architecture enables to improve the efficiency when cheap but inaccurate classifiers are followed by expensive but accurate classifiers.

Table 1: Strategies in multiple classifier systems. Adapted from [35].

Method	**Architecture**	**Level**	**Train.**	**Adapt.**	**Comments**
Class set reduction	Serial/Parallel	Rank/Conf.	Yes	No	Efficient
Voting, AND/OR	Parallel	Abstract	No	No	Assumes independency
Associative switch	Parallel	Abstract	Yes	Yes	Explores local expertise
Borda count	Parallel	Rank	Yes	No	Converts ranks to confidences
Logistic regression	Parallel	Rank/Conf.	Yes	No	Converts ranks to confidences
Dempster-Shafer	Parallel	Rank/Conf.	Yes	No	Fuses non-probabilistic scores
Prod, min, max	Parallel	Confidence	No	No	Assumes independency
Sum, median	Parallel	Confidence	No	No	Assumes independency; robust
Gen. Ensemble	Parallel	Confidence	Yes	No	Considers error correlations
Stacking	Parallel	Confidence	Yes	No	Exploits scarcity in data
Fuzzy Integrals	Parallel	Confidence	Yes	No	Fuses non-probabilistic scores
Bagging	Parallel	Confidence	Yes	No	Needs many classifiers
Random subspace	Parallel	Confidence	Yes	No	Needs many classifiers
Adaptive weighting	Parallel	Confidence	Yes	Yes	Explores local expertise
MLE	Parallel	Confidence	Yes	Yes	Explores local expertise
Boosting	Parallel/Hier.	Abstract	Yes	No	Needs many classifiers
Neural tree	Hierarchical	Confidence	Yes	No	Handles many classes
Hierarchical MLE	Hierarchical	Confidence	Yes	Yes	Explores local expertise

In the parallel architecture all classifiers are invoked independently and their outputs are combined. Most methods in the literature belong to this category, which can be further divided into two classes: 1) selection, and 2) fusion. In classifier selection, the different individual systems are considered "experts" in local regions of the feature space. The combination gives then more importance to the classifier closest to the input pattern in terms of area of expertise [32, 34]. On the other hand, classifier fusion assumes that all the classifiers are trained and their expertise combined over the whole feature space [30].

Some well-known combination strategies in multiple classifier systems are compared in Table 1 based on the previous properties.

2.1. Parallel classifier combination

Multiple classifier outputs are usually made comparable by mapping them to the $[0, 1]$ interval. This score normalization step will be detailed in the case of multimodal authentication in Sect. 3.2.3. For some classifiers, these normalized output scores can be considered *a posteriori* probabilities for the classes. Assuming further restrictions, e.g., that the individual classifiers use mutually independent subsets of features (which is realistic in the case of multimodal biometrics), fusion can be reduced to simple operators such as product or average. Kittler et al. [1] followed this approach in a probabilistic Bayesian framework and provided an example of multimodal biometric authentication fusing speech, frontal and profile images modalities. Considering M classifiers, C classes, and a given pattern Z that generates the feature vector B_j for classifier j, the classifiers are considered to give the *a posteriori* probability for each class ω_c, $c = 1, \ldots, C$: $P(\omega_c|B_j)$. Several ways to implement the fusion of the classifiers are then obtained based on the Bayes theorem and certain hypothesis:

Product Rule. Assign $Z \to \omega_c$ if

$$P^{(1-M)}(\omega_c) \prod_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \left[P^{(1-M)}(\omega_r) \prod_{j=1}^{M} P(\omega_r|B_j) \right]. \quad (1)$$

Sum Rule. Assign $Z \to \omega_c$ if

$$(1-M)P(\omega_c) + \sum_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \left[(1-M)P(\omega_r) + \sum_{j=1}^{M} P(\omega_r|B_j) \right]. \quad (2)$$

Max Rule. Assign $Z \to \omega_c$ if

$$\max_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \max_{j=1}^{M} P(\omega_r|B_j). \quad (3)$$

Min Rule. Assign $Z \to \omega_c$ if

$$\min_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \min_{j=1}^{M} P(\omega_r|B_j). \quad (4)$$

Median Rule. Assign $Z \to \omega_c$ if

$$\operatorname*{med}_{j=1}^{M} P(\omega_c|B_j) = \max_{r=1}^{C} \operatorname*{med}_{j=1}^{M} P(\omega_r|B_j). \quad (5)$$

Majority Vote Rule. In this case the combination is not at score level but at decision level. The *a posteriori* probabilities are thresholded to produce

$$\Delta_{rj} = \begin{cases} 1 & \text{if} \quad P(\omega_r|B_j) = \max_{c=1}^{C} P(\omega_c|B_j) \\ 0 & \text{otherwise} \end{cases} . \tag{6}$$

The majority vote rule then assigns $Z \to \omega_c$ if

$$\sum_{j=1}^{M} \Delta_{cj} = \max_{r=1}^{C} \sum_{j=1}^{M} \Delta_{rj}. \tag{7}$$

The product rule is obtained from the assumption of statistical independence of the different representations B_j with $j = 1, \ldots, M$. The sum rule is obtained further assuming that the *a posteriori* probabilities computed by the classifiers do not deviate much from the *a priori* probabilities, which is the case in a noisy scenario. The remaining rules are obtained by approximating or bounding the *a posteriori* probabilities. The sum rule outperformed the remainder in the experimental comparison. This was explained by a theoretical analysis of its robustness to the estimation errors of $P(c|B_j)$ made by the individual classifiers. Subsequent works have also reported comparative studies between these simple fusion rules [36, 37, 38].

Another paradigm for parallel classifier fusion is based on considering the combination stage as a second-level pattern recognition problem [4]. In this case the outputs from the different classifiers are considered as a new feature vector which is the input to a second-level classifier. The methods specially developed for multiple classifier combination (some of them summarized in Table 1), can therefore be extended with any of the large number of classifiers available from the literature.

2.2. Theoretical underpinnings in MCS

A large number of experimental studies have demonstrated the benefits of classifier combination [25]. However, very few works have provided some insight into the theoretical explanations.

One preliminary yet rigorous theory for classifier combination was developed by [39]. Another theoretical analysis of classifier combination was presented by [40], which is based on the well-known bias/variance dilemma [41].

Theoretical developments in multiple classifiers systems under severe restrictions usually assume linearly combined classifiers [42, 43]. As presented in previous section, another more general theoretical framework was presented in [1], who concluded that the weighted average combination is the most robust technique among the non-trained fusion rules evaluated. This result is also corroborated by the theoretical explanation by [44] for the effectiveness of the weighted average.

In the particular case of score fusion for biometric authentication, one of the very few works providing some theoretical insight was described by [45]. This study assumed Gaussian distributions of client and impostor scores and used a theoretical model called Variance Reduction-Equal Error rate (VR-EER). A number of findings linking the correlation and variance of base experts to the performance improvement of score fusion were then obtained.

Although the existence of these theoretical underpinnings, and the success of practical algorithms for classifier fusion, the problem of classifier combination is very complex and most aspects of a general theory still beg explanation [1]. Some of these not well known aspects include: relation between dimensionality expansion (multiple experts) and dimensionality reduction (expert combination), effect of individual expert error distribution on the choice of a combination strategy, etc. Furthermore, a number of practical multiple classifier approaches are either sequential or based on special rules for handling exceptions and rejections, which makes difficult the theoretical advance in this field.

3. MCS in multimodal biometrics

Multimodal biometric authentication can be seen as a two-class multiple classifier combination problem (either client or impostor). As such, most of the categories presented in Sect. 2 for general multiple classifier systems also apply here with some specificities. In particular, a biometric system is usually divided into four modules: 1) the sensor acquires the biometric data, 2) the feature extraction module process the biometric data in order to obtain a compact yet discriminative representation of the input biometric data, 3) the matching module compares input feature vectors to stored templates resulting in matching scores, and 4) the decision module releases an identification or verification decision based on the matching scores. Considering this architecture of biometric systems based on four modules, we adhere to

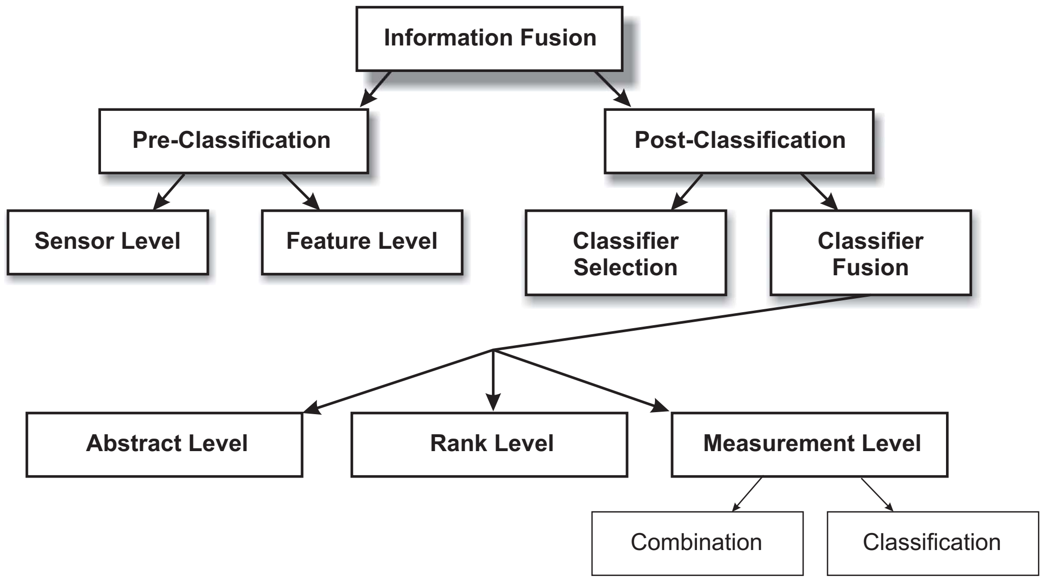

Figure 2: Approaches to information fusion in multimodal biometric authentication. Adapted from Jain et al. (2005) [46].

the taxonomy described in [46] to outline the state-of-the-art in multimodal biometric fusion. This taxonomy is sketched in Fig. 2.

3.1. Pre-classification fusion

Before classification/matching, integration of information can be done either at sensor level or feature level.

In *sensor level fusion*, raw data from the sensors are combined. One example is the combination of several cameras in face verification [47].

Feature level fusion refers to the combination of different feature vectors, obtained either with different sensors or by applying different feature extraction algorithms to the same data. Two simple feature fusion schemes are: 1) weighting, when the feature vectors are homogeneous, and 2) concatenation, when the feature vectors are non-homogeneous. Experiments for homogeneous and non-homogeneous feature level fusion with face and hand modalities were reported in [48].

3.2. Post-classification fusion

Approaches for combining information after the matching can be divided into classifier selection and classifier fusion. In the first category, the result is based only on the classifier most likely to give the correct decision for the input pattern. Classifier fusion can be further divided depending on the information to be combined: decisions, ranks, or matching scores.

Abstract or *decision level fusion* refers to the combination of decisions already taken by the individual biometric systems. Examples include: majority voting, weighted voting based on Dempster-Shafer theory [30], AND rule, OR rule, etc.

Rank level fusion take place when the individual systems provide a set of possible matches ranked according to confidence. This approach only makes sense in biometric identification, where a number of comparisons between each input pattern and the stored templates in the database are carried out. One example of rank level fusion is Borda count, which is based on the sum of ranks provided by the individual classifiers [49].

Score level fusion, also denoted as *measurement* or *confidence level fusion*, refers to the combination of matching scores provided by the different classifiers. In the context of biometric authentication, score level fusion can be classified into two categories: combination and classification. In the combination approach the input matching scores are normalized into the same range and then combined to obtain a scalar fused score. In the classification approach the matching scores are considered as input features for a second-level pattern classification problem between two classes, either client or impostor.

3.2.1. Combination approach

Combination approaches include: product, sum, max, min, median, and majority vote rules as described in Eqs. (1) to (7).

In the case of multimodal biometric authentication there are only two classes ($C = 2$): ω_0 = impostor and ω_1 = client. Let us also assume that the output similarity matching scores s_j from each system $j = 1, \ldots, M$ are normalized into x_j in order to have $x_j \approx P(\omega_1|B_j)$. The *a posteriori* probabilities for the impostor class are then $P(\omega_0|B_j) = 1 - x_j$. Under these common assumptions in multimodal biometric authentication, the classification rules in Eqs. (1) to (7) are simplified significantly. As an example, the sum rule in Eq. (2) is based on the evaluation of

$$(1-M)P(\omega_1) + \sum_{j=1}^{M} x_j > (1-M)P(\omega_0) + \sum_{j=1}^{M} (1-x_j), \tag{8}$$

which is equivalent to evaluating

$$y = \sum_{j=1}^{M} x_j > \frac{(1-M)P(\omega_0) - (1-M)P(\omega_1) + M}{2} = \text{Decision Threshold.} \tag{9}$$

This last result indicates that the general sum rule for combining classifiers reduces to simple matching score sum plus a decision based on a threshold. This decision threshold depends on the number of systems M and the *a priori* probabilities of client and impostor classes. The remaining rules can be similarly demonstrated to reduce to simple product, max, min, and median of matching scores plus a decision threshold. Variants including weighting parameters for each system can be also found in the literature [50]

$$\sum_{j=1}^{M} w_j x_j > \text{Decision Threshold.} \tag{10}$$

The parameters w_j can be computed heuristically, by exhaustive search in order to minimize certain error criterion on a training set, or by using a trained approach based on linear classifiers.

The previous rules assume that the output matching scores from the individual systems s_j have been mapped to *a posteriori* probabilities x_j, which by no means is a straightforward task and in most cases is not realistic. This issue is considered in more detail in Sect. 3.2.3.

Another theoretical framework which does not rely on the assumption of posterior probabilities released by the individual systems was developed by [51]. This work used Bayesian statistics to estimate the accuracy of individual classifiers during the fusion process. In brief, this Expert Conciliation approach results in a combination function based on weighted average of similarity scores x_j

$$y = \begin{cases} \sum_{j=1}^{M} w_j^{\mathcal{C}} x_j + w_0^{\mathcal{C}} & \text{if} \quad \left| 1 - \sum_{j=1}^{M} w_j^{\mathcal{C}} x_j + w_0^{\mathcal{C}} \right| < \left| \sum_{j=1}^{M} w_j^{\mathcal{I}} x_j + w_0^{\mathcal{I}} \right| \\ \sum_{j=1}^{M} w_j^{\mathcal{I}} x_j + w_0^{\mathcal{I}} & \text{otherwise} \end{cases}, \tag{11}$$

where the superindexes $\mathcal{C}$ and $\mathcal{I}$ denote parameters computed over a training set of client and impostor scores, respectively. Because this method is not built on the assumption of scores matching *a posteriori* probabilities, this combination approach does not rely so heavily on score normalization as the simple rules mentioned before.

Note that the combination approaches mentioned in this section are either fixed or trained. Simple rules such as product, sum, or max are fixed, although they rely on score normalization which may be subject to training. On the other hand, the Expert Conciliation scheme in Eq. (11) is a trained fusion approach.

As in every pattern recognition problem, the success of fixed rules depends heavily on the prior assumptions. On the other hand, the success of trained approaches relies heavily on the amount and representativeness of the training data. This tradeoff can be used to explain the contradictory results obtained in a number of works when comparing fixed to trained approaches, [52, 53]. In general, the success of a trained fusion scheme will depend on the conditions of the problem at hand including the prior information and the amount of training data [54].

3.2.2. Classification approach

In this category of methods, the normalized matching scores x_j, $j = 1, \ldots, M$ are joined together in a feature vector $[x_1, \ldots, x_M]^T$, which is the input to a two-class pattern classification problem, either client or impostor. Although some classification methods may work better when the input features are in the same range, the classification approach to fusion does not necessarily rely on score normalization, so we can assume either $x_j = s_j$ or a basic fixed score normalization just to make homogeneous the score ranges between different systems.

One early study using the classification approach in multimodal biometrics was reported in [55]. This pioneer work combined face (3 classifiers) and voice (2 classifiers) by using various forms of rank and measurement level fusion, including a Neural Network.

Chatzis et al. [56] combined in different ways five different unimodal experts, four for face and one for speech authentication. Experiments were performed by considering repeatedly each person as an impostor and the remaining persons as clients for every shot, with four shots per person. Fusion methods used were the following: OR and AND logical operators on thresholded scores, k-means algorithm, fuzzy k-means algorithm, fuzzy vector quantization algorithm, fuzzy k-means for fuzzy data, fuzzy vector quantization for fuzzy data, and median radial basis function network. For algorithms which operate on fuzzy data, data was fuzzified by quality measures of experts' opinions. This is one of the first published works that used quality measures in the framework of multimodal biometric fusion.

Verlinde et al. [57] followed the classification approach to fusion and compared a number of pattern classification techniques combining face profile, frontal face, and voice. The results sorted by relative decreasing performance were the following: Logistic Regression, Maximum a Posteriori, k-Nearest Neighbors, Multilayer Perceptrons, Binary Decision Trees, Maximum Likelihood, Quadratic Classifiers and Linear Classifiers. In a subsequent contribution [58], the paradigm of Support Vector Machines (SVM) was compared with all the above-mentioned techniques on the same experimental scenario, outperforming all of them. This is corroborated by other comparative studies [59], which favored the SVM approach over Neural Networks and Decision Trees. The comparisons were only based on recognition error rates. Therefore the comparative results should be taken with care, as other important factors may be considered in practical implementations, namely: ease of training, ease of implementation, scalability, etc.

Bengio et al. [60] performed fusion of two experts, face verification based on Neural Networks and voice verification based on Gaussian Mixture Models by using three different fusion algorithms: Multi-Layer Perceptrons (MLP), Support Vector Machines (SVM) and Bayes Classifiers using Gaussian Mixture Models (GMM) as density estimators. They compared the performance of each of these methods with and without estimation of confidence of unimodal scores. Intuitively, knowledge of confidence measures on these scores should help in the weighting process, i.e., if one multimodal system produces scores not very precisely, its score should be given less weight. Thus, they proposed and compared three methods to estimate a measure of confidence over a score. The first method is based on Gaussian hypothesis of the score distribution. The second method estimates the confidence by using a resampling technique based on groups of training scores. The third method is based on the adequacy of the trained models to explain the input biometric data. The conclusion of this study is that some confidence measures were able to enhance the fusion performance, but not systematically. In this study the confidence measures were obtained directly either from the available training scores or from parameters of the trained models, and not from the quality of the input biometric signals.

Roli et al. [61] estimated the performance of classifier ensembles consisting of two to eight different experts. Experts' opinions were combined by using five fixed and two trained fusion rules. Fixed rules included: sum, majority vote, and order statistics operators such as min, med and max. Trained rules included: weighted average, and Behavior Knowledge Space method. They concluded that it is better to combine the most complementary experts rather than the best performing ones. They also concluded that, in real applications, the poor quality and/or the limited size of the training set "can quickly cancel the theoretical advantages of trained rules". Among fixed rules, the vote majority rule exhibited good performance.

Ross and Jain [53] compared the performance of weighted sum, Decision Tree and Linear Discriminant Classifier for the fusion of face, fingerprint and hand geometry modalities. By using simple fixed score normalization, sum rule outperformed both Decision Tree and Linear Discriminant Classifiers.

In Table 2 we summarize some of the mentioned and other representative works in multimodal biometric fusion published in the last 25 years.

3.2.3. Score normalization

In general, the similarity matching scores s_j can be modelled as [46]

$$s_j = f[P(\omega_1|B_j)] + \eta(B_j), \tag{12}$$

where f is a monotonic function and η is the error made in the estimation of the *a posteriori* probability by the individual system j. This error can be due to noise in the input biometric signals or errors in the feature extraction or matching.

A number of works have focused on mapping output similarity scores s_j to *a posteriori* probabilities $P(\omega_1|B_j)$ by using different assumptions. Most of them assume $\eta(B_j) = 0$ in Eq. (12) and particular distributions for the similarity scores. Snelick et al. [81] assumed the conditional densities $P(s_j|\omega_0)$ and $P(s_j|\omega_1)$ to be Gaussian. A more general assumption was developed in [82] by using non-parametric density estimation based on Parzen Windows.

Either because of the unrealistic assumptions, or because of problems with density estimation on a finite training set, the prevalent method in the combination approach is not to map scores to probabilities but just to transform them into a common domain by using an operational technique for *score normalization* [46]. These techniques can be either *fixed* or *adaptive*. The most common techniques for fixed score normalization are:

Min-max. The matching scores s are normalized according to

Table 2: Summary of works on multimodal biometrics. M denotes the total number of classifiers combined. Architecture is either Serial or Parallel. Level is either Rank or Confidence. Performance gain over the best single classifier is given for IDentification or VERification either as FR@FA pair, EER or Total Error TE=FR+FA (in %). Adapted from [35].

Work	**Modalities**	**M**	**Arch.**	**Level**	**Gain**
Brunelli and Falavigna (1995) [55]	Speaker, face	5	P	C	ID:17→2 (TE)
Duc et al. (1997) [62]	Speaker, face	2	P	C	VER:6.7→0.5 (TE)
Kittler et al. (1998) [1]	Speaker, face	3	P	C	VER:1.4→0.7 (EER)
Hong and Jain (1998) [63]	Face, fingerprint	2	S	R/C	ID:6.9→4.5 (FR@0.1%FA)
Jain et al. (1999) [64]	Speaker, face, finger	3	P	C	VER:15→3 (FR@0.1%FA)
Ben-Yacoub et al. (1999) [59]	Speaker, face	3	P	C	VER:4→0.5 (EER)
Choudhury et al. (1999) [65]	Speaker, face	3	P	C	ID:16.5→6.5 (TE)
Chatzis et al. (1999) [56]	Speaker, face	4	P	C	ID:6.7→1.07 (TE)
Verlinde et al. (2000) [57]	Speaker, face	3	P	C	VER:3.7→0.1 (TE)
Ross and Jain (2003) [53]	Face, finger, hand	3	P	C	VER:16→2 (FR@0.1%FA)
Kumar and Zhang (2003) [66]	Face, palmprint	2	P	C	VER:3.6→0.8 (EER)
Wang et al. (2004) [67]	Speaker, finger	2	P	C	VER:2→0.7 (EER)
Fierrez et al. (2005) [16]	Signature, finger	2	P	C	VER:3.5→1.1 (EER)
Poh and Bengio (2006) [68]	Speaker, face	8	P	C	VER:2.2→0.7 (TE)
Sim et al. (2007) [69]	Face, finger	2	P	C	f(data accross time)
Fronthaler et al. (2008) [70]	Finger	3	P,S	C	f(architecture,quality)
Poh et al. (2009) [71]	Face, iris, finger	24	P,S	C	f(missing data,device)
Ortega et al. (2010) [72]	Face, iris, finger, signature	n/a	P,S	C	f(acquisition scenario)
Dantcheva et al. (2011) [73]	Face, body	9	P	C	f(soft biometrics)
Biggio et al. (2012) [74]	Face, finger	2	P	C	f(attack type)
Shekhar et al. (2014) [75]	Face, iris, finger	6	P	C	f(noise,occlusion,disguise)
Tome et al. (2014) [76]	Face, body	23	P	C	f(distance)
Meng et al. (2015) [77]	Overview	-	-	-	f(architecture)
Hadid et al. (2015) [78]	Overview	-	-	-	f(attack type)
Singh et al. (2016) [79]	Overview	-	-	-	f(recent methods)
Lumini and Nanni (2017) [15]	Overview	-	-	-	Intro + Case study
Gomez et al. (2017) [80]	Signature, fingerprint	2	P	C	f(security,complexity)

$$x = \frac{s - \min}{\max - \min}, \tag{13}$$

where the maximum and minimum are computed from a given set of training scores. This normalization method is specially prone to errors due to outliers.

Z-score. The matching scores s are normalized with

$$x = \frac{s - \mu}{\sigma}, \tag{14}$$

where μ and σ are respectively the arithmetic mean and standard deviation of a given set of training scores.

Exponential functions. This include various forms of exponentials [17]

$$x = c_1 \exp(c_2 s) + c_3, \tag{15}$$

sigmoids [83, 81]

$$x = \frac{c_1}{1 + \exp(c_2 s + c_3)} + c_4, \tag{16}$$

or hyperbolic functions [46]

$$x = c_1 \tanh(c_2 s + c_3) + c_4, \tag{17}$$

where c_1 to c_4 are parameters. As demonstrated by [46], exponential-based score normalization is more robust and efficient than min-max and z-score, where robust refers to insensitivity to the presence of outliers, and efficiency refers to the proximity of the obtained estimate to the optimal estimate when the distribution of the data is known.

After mapping the matching scores s_j to a common domain x_j, simple combination rules as in Eq. (9) are then usually applied.

Adaptive score normalization techniques modify the score normalization functions depending on the context, e.g., for each user [84], or depending on the input quality [85].

4. Conclusions

In the present paper, Part 1 in a series of two papers, we have first provided a brief introduction to Multiple Classifier Systems (MCS) including basic nomenclature, architecture, and key elements. Our main focus has been into the fundamentals of MCS, providing pointers where the reader may find detailed descriptions of established MCS algorithms.

The presentation has been kept as general as possible, in order to be useful for any pattern recognition application where MCS may be applied, e.g.: biometrics person recognition [19], video surveillance [20], speech technologies [21], biomedical applications [86], human-computer interaction [22], data analytics [23], or recommender systems [24].

We have then overviewed the application of MCS to the particular field of multimodal biometric person authentication in the last 25 years [19], including general descriptions of main MCS elements, methods, and algorithms. Our presentation has been kept general with a generic mathematical formulation, in order to facilitate the export of experiences and methods to other information fusion problems like the above mentioned. Based on our mathematical presentation, we have also discussed the main theoretical advances and current state towards a comprehensive theory in the field of MCS.

Based on the theory and framework introduced here, in the companion paper we develop in more technical detail recent trends and developments in MCS from multimodal biometrics that incorporate context information in an adaptive way. These new MCS architectures exploit input quality measures [16] and pattern-specific particularities that move apart from general population statistics [17], resulting in robust multimodal biometric systems. Similarly as in the present paper, methods in the companion paper are introduced in a general way so they can be applied to other information fusion problems as well.

Finally, also in the companion paper, we also discuss open challenges in biometrics in which MCS may play a key role.

5. Acknowledgements

This work was funded by projects CogniMetrics (TEC2015-70627-R) from MINECO/FEDER and RiskTrakc (JUST-2015-JCOO-AG-1). Part of this work was conducted during a research visit of J.F. to Prof. Ludmila Kuncheva at Bangor University (UK) with STSM funding from COST CA16101 (MULTI-FORESEE). Author J.F. want to thank Prof. Kuncheva for fruitful discussions during his visit.

References

[1] J. Kittler, M. Hatef, R. Duin, J. Matas, On combining classifiers, IEEE Trans. on Pattern Anal. and Machine Intell. 20 (3) (1998) 226–239.

[2] J. Kittler, F. Roli (Eds.), Multiple Classifier Systems, First International Workshop, Vol. 1857 of Lecture Notes on Computer Science, Springer, 2000.

[3] F. Schwenker, J. Kittler, F. Roli (Eds.), Multiple Classifier Systems, 12th International Workshop, Vol. 9132 of Lecture Notes on Computer Science, Springer, 2015.

[4] L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Wiley, 2014.

[5] H. M. Gomes, J. P. Barddal, F. Enembreck, A. Bifet, A survey on ensemble learning for data stream classification, ACM Comput. Surv. 50 (2) (2017) 23:1–23:36.

[6] M. Re, G. Valentini, Advances in Machine Learning and Data Mining for Astronomy, Chapman and Hall/CRC, 2012, Ch. Ensemble Learning.

[7] A. A. Ross, K. Nandakumar, A. K. Jain, Handbook of Multibiometrics, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[8] A. Criminisi, J. Shotton, Decision Forests for Computer Vision and Medical Image Analysis, Springer, 2013.

[9] R. Polikar, Ensemble based systems in decision making, IEEE Circuits and Systems Magazine 6 (3) (2006) 21–45.

[10] S. Tulyakov, S. Jaeger, V. Govindaraju, D. Doermann, Machine Learning in Document Analysis and Recognition, Springer Berlin Heidelberg, 2008, Ch. Review of Classifier Combination Methods, pp. 361–386.

[11] M. Wozniak, M. Grana, E. Corchado, A survey of multiple classifier systems as hybrid systems, Information Fusion 16 (2014) 3 – 17.

[12] G. Brown, Encyclopedia of Machine Learning, Springer US, Boston, MA, 2010, Ch. Ensemble Learning, pp. 312–320.

[13] F. Gargiulo, C. Mazzariello, C. Sansone, Handbook on Neural Information Processing, Springer Berlin Heidelberg, 2013, Ch. Multiple Classifier Systems: Theory, Applications and Tools, pp. 335–378.

[14] P. Viola, M. J. Jones, Robust real-time face detection, Int. J. Comput. Vision 57 (2) (2004) 137–154.

[15] A. Lumini, L. Nanni, Overview of the combination of biometric matchers, Information Fusion 33 (2017) 71 – 85.

[16] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, J. Bigun, Discriminative multimodal biometric authentication based on quality measures, Pattern Recognition 38 (5) (2005) 777–779.

[17] J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, J. Gonzalez-Rodriguez, Adapted user-dependent multimodal biometric authentication exploiting general information, Pattern Recognition Letters 26 (16) (2005) 2628–2639.

[18] L. Snidaro, J. Garca, J. Llinas, Context-based information fusion: A survey and discussion, Information Fusion 25 (2015) 16 – 31.

[19] A. K. Jain, K. Nandakumar, A. Ross, 50 years of biometric research, Pattern Recogn. Lett. 79 (2016) 80–105.

[20] A. Garcia-Martin, J. M. Martinez, People detection in surveillance: classification and evaluation, IET Computer Vision 9 (2015) 779–788(9).

[21] I. Lopez-Moreno, J. Gonzalez-Dominguez, D. Martinez, O. Plchot, J. Gonzalez-Rodriguez, P. J. Moreno, On the use of deep feedforward neural networks for automatic language identification, Computer Speech and Language 40 (2016) 46 – 59.

[22] D. Rozado, T. Moreno, J. S. Agustin, F. B. Rodriguez, P. Varona, Controlling a smartphone using gaze gestures as the input mechanism, Human-Computer Interaction 30 (1) (2015) 34–63.

[23] G. Bello-Orgaz, J. J. Jung, D. Camacho, Social big data: Recent achievements and new challenges, Information Fusion 28 (2016) 45 – 59.

[24] P. Castells, N. J. Hurley, S. Vargas, Recommender Systems Handbook, Springer US, 2015, Ch. Novelty and Diversity in Recommender Systems, pp. 881–918.

[25] A. K. Jain, R. P. W. Duin, J. Mao, Statistical pattern recognition: A review, IEEE Trans. on Pattern Anal. and Machine Intell. 22 (1) (2000) 4–37.

[26] J. F. Diez-Pastor, J. J. Rodriguez, C. I. Garcia-Osorio, L. I. Kuncheva, Diversity techniques improve the performance of the best imbalance learning ensembles, Information Sciences 325 (2015) 98 – 117.

[27] D. Wolpert, Stacked generalization, Neural Networks 5 (1992) 241–259.

[28] L. Breiman, Bagging predictors, Machine Learning 24 (2) (1996) 123–140.

[29] R. Shapire, The strength of weak learnability, Machine Learning 5 (1990) 197–227.

[30] L. Xu, A. Kryzak, C. Suen, Methods of combining multiple classifiers and their application to handwritten recognition, IEEE Trans. on Systems, Man and Cybernetics 22 (3) (1992) 418–435.

[31] V. Tresp, M. Taniguchi, Combining estimators using non-constant weighting functions, in: Advances in Neural Information Processing Systems, NIPS, Vol. 7, MIT Press, 1995.

[32] R. Jacobs, M. Jordan, S. Nowlan, G. Hinton, Adaptive mixtures of local experts, Neural Comput. 3 (1991) 79–87.

[33] M. I. Jordan, R. A. Jacobs, Hierarchical mixtures of experts and the EM algorithm, Neural Computation 6 (1994) 181–214.

[34] E. Alpaydin, M. Jordan, Local linear perceptrons for classification, IEEE Trans. Neural Networks 7 (3) (1996) 788–792.

[35] D. Maltoni, D. Maio, A. Jain, S. Prabhakar, Handbook of Fingerprint Recognition, Springer, 2009, Ch. Biometric Fusion, pp. 330–339.

[36] F. Alkoot, J. Kittler, Improving the performance of the product fusion strategy, in: Proc. of Intl. Conf. on Pattern Recognition, ICPR, Vol. 2, 2000, pp. 164–167.

[37] L. I. Kuncheva, A theoretical study on six classifier fusion strategies, IEEE Trans. on Pattern Analysis and Machine Intelligence 24 (4) (2002) 281–286.

[38] J. Kittler, F. M. Alkoot, Sum versus vote fusion in multiple classifier systems, IEEE Trans. on Pattern Analysis and Machine Intelligence 25 (1) (2003) 110–115.

[39] R. Kleinberg, Stochastic discrimination, Annals of Math. and Artificial Intelligence 1 (1990) 207–239.

[40] A. Krogh, J. Vedelsby, Neural network ensembles, cross validation, and active learning, in: Advances in Neural Information Processing Systems, NIPS, Vol. 7, MIT Press, 1995.

[41] S. Geman, E. Bienenstock, R. Doursat, Neural networks and the bias/variance dilemma, Neural Computation 4 (1) (1992) 1–58.

[42] K. Tumer, J. Ghosh, Analysis of decision boundaries in linearly combined neural classifiers, Pattern Recognition 29 (1996) 341–348.

[43] G. Fumera, F. Roli, A theoretical and experimental analysis of linear combiners for multiple classifier systems, IEEE Trans. on Pattern Analysis and Machine Intelligence 27 (6) (2005) 942–956.

[44] R. Shapire, Y. Freund, P. Bartlett, W. Lee, Boosting the margin: A new explanation for the effectiveness of voting methods, The Annals of Statistics 26 (5) (1998) 1651–1686.

[45] N. Poh, S. Bengio, How do correlation and variance of base classifiers affect fusion in biometric authentication tasks?, IEEE Trans. on Signal Processing 53 (11) (2005) 4384–4396.

[46] A. K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognition 38 (2005) 2270–2285.

[47] K. Chang, K. Bowyer, P. Flynn, An evaluation of multimodal 2D+3D face biometrics, IEEE Trans. on Pattern Analysis and Machine Intelligence 27 (2005) 619–624.

[48] A. Ross, R. Govindarajan, Feature level fusion using hand and face biometrics, in: Proc. of Intl. Conf. on Biometric Technologies for Human Identification, BTHI, Vol. 5779, Proc. of SPIE, 2005, pp. 196–204.

[49] T. Ho, J. Hull, S. Srihari, Decision combination in multiple classifier systems, IEEE Trans. on Pattern Analysis and Machine Intelligence 16 (1) (1994) 66–75.

[50] A. K. Jain, A. Ross, Learning user-specific parameters in a multibiometric system, in: Proc. of IEEE Intl. Conf. on Image Processing, ICIP, Vol. 1, 2002, pp. 57–60.

[51] E. S. Bigun, J. Bigun, B. Duc, S. Fischer, Expert conciliation for multi modal person authentication systems by Bayesian statistics, in: J. Bigun, G. Chollet, G. Borgefors (Eds.), Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA, Springer LNCS-1206, 1997, pp. 291–300.

[52] F. Roli, G. Fumera, J. Kittler, Fixed and trained combiners for fusion of imbalanced pattern classifiers, in: Proc. of the Intl. Conf. on Information Fusion, FUSION, 2002, pp. 278–284.

[53] A. Ross, A. K. Jain, Information fusion in biometrics, Pattern Recognition Letters 24 (13) (2003) 2115–2125.

[54] R. P. W. Duin, The combining classifier: to train or not to train?, in: Proc. of the IAPR Intl. Conf. on Pattern Recognition, ICPR, IEEE CS Press, 2002, pp. 765–770.

[55] R. Brunelli, D. Falavigna, Person identification using multiple cues, IEEE Trans. on Pattern Anal. and Machine Intell. 17 (10) (1995) 955–966.

[56] V. Chatzis, A. G. Bors, I. Pitas, Multimodal decision-level fusion for person authentication, IEEE Trans. on System, Man, and Cybernetics, part A 29 (6) (1999) 674–680.

[57] P. Verlinde, G. Chollet, M. Acheroy, Multi-modal identity verification using expert fusion, Information Fusion 1 (1) (2000) 17–33.

[58] B. Gutschoven, P. Verlinde, Multi-modal identity verification using Support Vector Machines (SVM), in: Proc. of Intl. Conf. on Information Fusion, FUSION, IEEE Press, 2000, pp. 3–8.

[59] S. Ben-Yacoub, Y. Abdeljaoued, E. Mayoraz, Fusion of face and speech data for person identity verification, IEEE Trans. on Neural Networks 10 (5) (1999) 1065–1074.

[60] S. Bengio, C. Marcel, S. Marcel, J. Mariethoz, Confidence measures for multimodal identity verification, Information Fusion 3 (4) (2002) 267–276.

[61] F. Roli, J. Kittler, G. Fumera, D. Muntoni, An experimental comparison of classifier fusion rules for multimodal personal identity verification systems, in: Proc. of Third Intl. Workshop on Multiple Classifier Systems, MCS, 2002, pp. 252–261.

[62] B. Duc, E. S. Bigun, J. Bigun, G. Maitre, S. Fischer, Fusion of audio and video information for multi modal person authentication, Pattern Recognition Letters 18 (1997) 835–843.

[63] L. Hong, A. K. Jain, Integrating faces and fingerprints for personal identification, IEEE Trans. on Pattern Anal. and Machine Intell. 20 (12) (1998) 1295–1307.

[64] A. K. Jain, L. Hong, Y. Kulkarni, A multimodal biometric system using fingerprint, face and speech, in: Proc. of Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA-99, 1999, pp. 182–187.

[65] T. Choudhury, B. Clarkson, T. Jebara, A. Pentland, Multimodal person recognition using unconstrained audio and video, in: Proc. of Intl. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA, 1999, pp. 176–181.

[66] A. Kumar, D. Zhang, Integrating palmprint with face for user authentication, in: Proc. of Workshop on Multimodal User Authentication, MMUA, 2003, pp. 107–112.

[67] Y. Wang, Y. Wang, T. Tan, Combining fingerprint and voice biometrics for identity verification: An experimental comparison, in: D. Zhang, A. K. Jain (Eds.), Proc. of Intl. Conf. on Biometric Authentication, ICBA, Springer LNCS-3072, 2004, pp. 663–670.

[68] N. Poh, S. Bengio, Database, protocol and tools for evaluating score-level fusion algorithms in biometric authentication, Pattern Recognition 39 (2) (2006) 223–233.

[69] T. Sim, S. Zhang, R. Janakiraman, S. Kumar, Continuous verification using multimodal biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 687–700.

[70] H. Fronthaler, K. Kollreider, J. Bigun, J. Fierrez, F. Alonso-Fernandez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, Fingerprint image quality estimation and its application to multi-algorithm verification, IEEE Trans. on Information Forensics and Security 3 (2) (2008) 331–338.

[71] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-Fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-Garcia, D. Maurer, A. A. Salah, T. Scheidat, C. Vielhauer, Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms, IEEE Trans on Information Forensics and Security 4 (4) (2009) 849–866.

[72] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. Freire, J. Gonzalez-Rodriguez, C.Garcia-Mateo, J.-L.Alba-Castro, E.Gonzalez-Agulla, E.Otero-Muras, S.Garcia-Salicetti, L.Allano, B.Ly-Van, B.Dorizzi, J.Kittler, T.Bourlai, N.Poh, F.Deravi, M.Ng, M.Fairhurst, J.Hennebert, A.Humm, M.Tistarelli, L.Brodo, J.Richiardi, A.Drygajlo, H.Ganster, F.M.Sukno, S.-K.Pavani, A.Frangi, L.Akarun, A.Savran, The multi-scenario multi-environment biosecure multimodal database (bmdb), IEEE Trans. on Pattern Analysis and Machine Intelligence 32 (6) (2010) 1097–1111.

[73] A. Dantcheva, C. Velardo, A. D'Angelo, J.-L. Dugelay, Bag of soft biometrics for person identification, Multimedia Tools and Applications 51 (2) (2011) 739–777.

[74] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, F. Roli, Security evaluation of biometric authentication systems under real spoofing attacks, IET Biometrics 1 (1) (2012) 11–24.

[75] S. Shekhar, V. M. Patel, N. M. Nasrabadi, R. Chellappa, Joint sparse representation for robust multimodal biometrics recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (1) (2014) 113–126.

[76] P. Tome, J. Fierrez, R. Vera-Rodriguez, M. Nixon, Soft biometrics and their application in person recognition at a distance, IEEE Transactions on Information Forensics and Security 9 (3) (2014) 464–475.

[77] W. Meng, D. S. Wong, S. Furnell, J. Zhou, Surveying the development of biometric user authentication on mobile phones, IEEE Communications Surveys Tutorials 17 (3) (2015) 1268–1293.

[78] A. Hadid, N. Evans, S. Marcel, J. Fierrez, Biometrics systems under spoofing attack: An evaluation methodology and lessons learned, IEEE Signal Processing Magazine 32 (5) (2015) 20–30.

[79] R. Singh, A. Ross, K. W. Bowyer, Special issue on information fusion in biometrics, Information Fusion 32 (Part B) (2016) 1 – 2.

[80] M. Gomez-Barrero, E. Maiorana, J. Galbally, P. Campisi, J. Fierrez, Multi-biometric template protection based on homomorphic encryption, Pattern Recognition 67 (2017) 149–163.

[81] R. Snelick, U. Uludag, A. Mink, M. Indovina, A. K. Jain, Large scale evaluation of multimodal biometric authentication using state-of-the-art systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 450–455.

[82] S. Prabhakar, A. Jain, Decision-level fusion in fingerprint verification, Pattern Recognition 35 (4) (2002) 861–874.

[83] R. Cappelli, D. Maio, D. Maltoni, A multi-classifier approach to fingerprint classification, Pattern Analysis and Applications 5 (2) (2002) 136–144.

[84] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, Target dependent score normalization techniques and their application to signature verification, IEEE Trans. on Systems, Man and Cybernetics, part C 35 (3) (2005) 418–425.

[85] N. Poh, J. Kittler, T. Bourlai, Quality-based score normalization with device qualitative information for multimodal biometric fusion, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 40 (3) (2010) 539–554.

[86] L. Nanni, C. Salvatore, A. Cerasa, I. Castiglioni, Combining multiple approaches for the early diagnosis of alzheimer's disease, Pattern Recognition Letters 84 (2016) 259 – 266.